

IR – PROJECT

M.Venkateswarlu

S20180010107

Project Details:

- ➔ Domain: Sports
- ➔ Algorithms Used: TF.IDF weightage, Cosine Similarity

Implementation:

- ➔ Read the Data Set
- ➔ I created Inverted Index by removing stop words and storing each unique word as a key, posting list as value for the collection of documents from scratch without any inbuilt libraries.
- ➔ I saved the inverted index result into Json file using Json package called dump.

Task-1:

Searching and retrieving the top 10 ranked documents.

- ➔ Input query is represented as vector with each dimension as each word.
- ➔ Now for calculating TF.IDF score we need term frequency and document frequency.
- ➔ For calculating term frequency I am checking the intersection between Input phrasal query terms and document terms, for the intersection terms only i am calculating the term frequency in that particular document.

- ➔ For calculating the document frequency I used Inverted Index as I stored posting list as dictionary value, simply size of Posting list corresponding to a term gives document frequency of that term in collection.
- ➔ I multiplied the TF.IDF scores obtained from between each common term between query and document.
- ➔ Then based on resultant TF.IDF score for each document I am retrieving the Top 10 Ranked documents.

-----Tasks-----

- 1.Searching and retrieving the top 10 ranked documents
- 2.Recommendations on Query Expansion based on previous searches
- 3.Exit

Enter the task number to perform

1

Enter a phrasal query Federer Overall GrandSlams

-----Document No----- 25777

Federer loses opening match at Shanghai Masters Tuesday, 13 October 2015 (3 minutes ago) SHANGHAI (AP) Defending champion Roger Federer lost his opening match at the Shanghai Masters on Tuesday, falling to Spanish qualifier Albert Ramos-Vinolas 7-6 (4), 2-6, 6-3. Roger Federer was upset in his opening match at the Shanghai Masters on Tuesday by Spanish qualifier Albert Ramos-Vinolas 7-6 (4), 2-6, 6-3. MyNorthwest.com 2 hours ago - Sports Former U.S. Open champion Marin Cilic fought off a match point before rallying to defeat local hope Wu Di of China 4-6, 6-3, 7-6 (4) in the opening round of the... MyNorthwest.com 22 hours ago - Sports Nick Kyrgios has enjoyed a straightforward opening to his Shanghai Masters campaign but things quickly get harder with a match against Kei Nishikori next. The Age 1 day ago - World You Might Like Other recent news in Sports Comments Twitter Damion Ackerly "Federer Loses Opening Match at Shanghai Masters" by THE ASSOCIATED PRESS via NYT <http://t.co/DODtuJmYxC> 35 seconds ago Regina Daily News Roger Federer loses opening match at Shanghai Masters: Defending champion Roger Federer lost his opening match at... <http://t.co/dCigsX4jLP> 2 minutes ago Worldwide News Roger Federer loses opening match at Shanghai Masters <http://t.co/W4070pb3CH> 2 minutes ago Empire Financial SA Federer loses opening match at Shanghai Masters <http://t.co/W30LHZSMFS> 2 minutes ago Faye Sarazin "Federer Loses Opening Match at Shanghai Masters" by THE ASSOCIATED PRESS via NYT <http://t.co/vn9HpYLIbo> 3 minutes ago The Truth 24 UK Federer loses opening match at Shanghai Masters <http://t.co/KeNlZoYXUB> 4 minutes ago

---TF-IDF:Score--> 13.71793022962541

This is the top one ranked document

Task-2:

User Recommendations on Query Expansion based on previous searches.

- ➔ Input query is represented as vector with each dimension as each word.
- ➔ I stored all previous query vectors in one list.
- ➔ Based on that cosine similarity Score between input phrasal query and previous queries, I am displaying the recommendations for that Query.

-----Tasks-----

- 1.Searching and retrieving the top 10 ranked documents
- 2.Recommendations on Query Expansion based on previous searches
- 3.Exit

```
Enter the task number to perform
2
Enter a phrasal query  Federer is
Recommendations

who federer
federer overall grandslams
```